



Démarches sémantiques de recherche d'information sur le Web

Olivier Corby, Catherine Faron Zucker, Isabelle Mirbel

► To cite this version:

Olivier Corby, Catherine Faron Zucker, Isabelle Mirbel. Démarches sémantiques de recherche d'information sur le Web. IC 2009: 20es Journées Francophones d'Ingénierie des Connaissances "Connaissance et communautés en ligne", May 2009, Tunisie. pp.289-300. hal-00380309

HAL Id: hal-00380309

<https://hal.science/hal-00380309>

Submitted on 30 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Démarches sémantiques de recherche d'information sur le Web

Olivier Corby¹, Catherine Faron-Zucker² et Isabelle Mirbel^{1,2}

¹EDELWEISS, INRIA Sophia Antipolis, Sophia Antipolis - France

²KEWI, I3S, Université de Nice Sophia - CNRS, Sophia Antipolis - France

olivier.corby@sophia.inria.fr,

catherine.faron-zucker@unice.fr, isabelle.mirbel@unice.fr

Type de communication : Recherche

Thèmes : Web sémantique

Résumé : L'examen de différents projets de recherche visant à supporter les activités des membres d'une communauté à l'aide d'une mémoire collective met en évidence l'intérêt de capitaliser les requêtes formulées à la mémoire et plus généralement les savoir-faire experts d'une communauté en matière de recherche d'information. L'enjeu est de donner les moyens aux membres d'une communauté de réutiliser et partager ces savoir-faire pour retrouver des informations précises et complètes par composition des résultats de requêtes sur différentes sources d'information. Dans cet article, nous proposons un modèle fondé sur les standards du web sémantique pour capitaliser, réutiliser et partager des séquences complexes de requêtes que nous appelons démarches de recherche. Notre modèle est le résultat d'une adaptation de la représentation intentionnelle de processus : nous explicitons les sous-buts qui gouvernent l'organisation d'une démarche de recherche et l'ordre selon lequel ces sous-buts doivent être satisfaits. Les démarches de recherche sont représentées en RDF et opérationnalisées par des requêtes représentées en SPARQL. L'instanciation d'une démarche repose sur la mise en oeuvre d'un mécanisme de chaînage arrière sur ces règles.

Mots-clés : Web sémantique, SPARQL, RDF(S), requêtes, démarches de recherche

1 Introduction

Les mémoires collectives permettent de supporter les activités de capitalisation, gestion et diffusion des connaissances au sein d'une communauté. Les ressources de la communauté y sont indexées par des annotations sémantiques qui explicitent et formalisent leur contenu informatif. Les membres d'une communauté exploitent alors leur mémoire communautaire en exprimant des requêtes leur permettant de rechercher des ressources pertinentes pour mener à bien leurs activités. La recherche des ressources de la mémoire qui répondent à une requête repose sur la manipulation formelle des annotations des ressources et peut être guidée par des ontologies du domaine.

L'examen de projets de recherche visant à assister les activités de communauté par une mémoire collective auxquels nous avons participé tels que le projet européen SevenPro (Cherfi *et al.*, 2008), les projets ANR e-WOK HUB (eWOKHUB Consortium, 2008) et Immunosearch (Kefi-Khelif *et al.*, 2008) ou le projet C3R (Yurchyshyna *et al.*, 2008) met en évidence le besoin de capitaliser les requêtes formulées par les utilisateurs dans une base idoine afin de permettre à leurs auteurs de les réutiliser, voire de les échanger avec les autres membres de la communauté. Plus généralement, la capitalisation des démarches de recherche d'information devient un véritable enjeu dans de nombreux domaines. En effet, des stratégies précises sont mises en oeuvre par les experts d'un domaine pour retrouver l'information nécessaire à leurs activités (Bhavnani *et al.*, 2003) et elles sont souvent difficiles à acquérir par des utilisateurs néophytes. Ces stratégies, de plus en plus critiques de par la spécialisation et la multiplication des bases de connaissances, sont rarement capitalisées et ne sont exploitées ni dans les outils de recherche comme Google ni dans les portails de domaine (Bhavnani *et al.*, 2003).

(Bhavnani *et al.*, 2003) proposent une approche pour expliciter les procédures critiques de recherche d'information dans le domaine médical à l'aide de ce qu'ils appellent des portails de stratégies de recherche («strategy hubs»). En partant d'un ensemble de questions types du domaine, ils définissent un ensemble de patrons représentant des procédures de recherche. Une procédure de recherche est représentée par un ensemble ordonné de sous-buts et pour chaque procédure de recherche, des liens vers les sources d'information pertinentes sont établis.

Les auteurs de (Buffereau *et al.*, 2003) proposent un environnement de navigation parmi des ressources Web qui distingue 3 niveaux de connaissances : (i) un niveau support de la connaissance qui rassemble les ressources du Web dans le domaine d'application, (ii) un niveau représentation de la connaissance qui rassemble les méta-données relatives aux ressources du niveau précédent et (iii) un niveau transmission de la connaissance qui propose des parcours (appelés «e-parcours») pour exploiter les ressources du Web au travers des méta-données qui leur sont associées. Les *e-parcours* sont composés d'étapes caractérisées par une intention ou un titre, un sujet et une illustration. Les illustrations sont les ressources Web relatives à l'étape du *e-parcours*.

Des exemples de procédures de recherche permettant une navigation dynamique existent également dans les systèmes d'apprentissage en ligne fondés sur les modèles du Web sémantique. Dans (Yessad *et al.*, 2008; Dehors & Faron-Zucker, 2006), nous modélisons l'approche pédagogique adoptée par un assemblage de requêtes paramétrées et les ressources dont les annotations répondent à ces requêtes composent dynamiquement des documents pédagogiques présentés à l'apprenant qui navigue dans le système. Dans nos travaux appliqués au domaine du bâtiment, nous avons explicité auprès des experts du domaine des procédures de vérification de conformité que nous représentons par un ordonnanceur des requêtes de conformité (Yurchyshyna *et al.*, 2008) auxquelles est soumise l'annotation de la maquette numérique d'un projet de construction à contrôler. L'ordonnancement dépend des classes auxquelles appartiennent les requêtes à mettre en oeuvre (types de textes réglementaires, types de bâtiments, parties de la maquette visée), c'est-à-dire que l'ordonnancement des classes de requêtes est fixe.

L'approche que nous proposons ici consiste en un modèle pour capitaliser, réutiliser et partager des requêtes de recherche d'information et pour les organiser en des démarches

de recherche formalisées également réutilisables et partageables. Comme dans les travaux cités plus haut, exceptés ceux de (Bhavnani *et al.*, 2003), notre approche repose sur les techniques et modèles du Web sémantique. Nous tirons parti des capacités d'inférence des ontologies capitalisant la connaissance d'un domaine d'application et nous utilisons le langage SPARQL qui permet de représenter des requêtes plus riches que les patrons proposés dans l'approche de (Bhavnani *et al.*, 2003). Comme dans (Buffereau *et al.*, 2003), nous proposons un modèle pour formaliser les buts et les sous-buts, qui sont exprimés en langage naturel dans (Bhavnani *et al.*, 2003). Enfin, notre proposition diffère de celle de (Buffereau *et al.*, 2003) dans le sens où elle permet la réutilisation et le partage d'étapes (i.e. de sous-buts) entre les procédures de recherche, alors que l'approche présentée dans (Buffereau *et al.*, 2003) positionne le partage uniquement au niveau des ressources Web. Précisément, nous nous intéressons à la construction et l'exploitation d'une base de requêtes rendant opérationnelles les étapes d'un scénario de recherche d'information que nous appelons démarche de recherche d'information. Cette base de connaissances peut être vue comme une mémoire épisodique dans laquelle les démarches de recherche sont construites dynamiquement en fonction du contexte.

La suite de l'article est organisée de la façon suivante. Dans la section 2 nous définissons la notion de démarche de recherche d'information et nous expliquons notre choix d'une modélisation intentionnelle de telles démarches. Dans la section 3, nous détaillons comment nous avons adapté un modèle intentionnel de représentation des processus au domaine du Web sémantique, au travers de la proposition d'une ontologie idoine permettant l'annotation de démarche de recherche d'information. Dans la section 4, nous présentons la façon dont les démarches de recherche sont mises en oeuvre à l'aide de règles.

2 Des démarches intentionnelles pour une compréhension globale d'une procédure de recherche

Nous définissons la notion de démarche de recherche d'information comme une séquence de recherches atomiques qui doivent être effectuées par un expert du domaine pour mener à bien une tâche ou un processus métier. Une démarche de recherche peut être vue comme un type particulier de processus métier constitué exclusivement d'activités de recherche d'information. Différents modèles de représentation des processus métiers ont été proposés dans la littérature (Nurcan & Edme, 2005).

Notre travail porte sur des moyens de capitaliser et d'explicitier des démarches de recherche permettant d'avoir une compréhension globale d'un sujet à partir de sources d'information dispersées dans différentes bases de connaissances et de données.

Pour favoriser le transfert de connaissances en matière de recherche d'information des experts vers les néophytes au sein d'une communauté, nous cherchons à nous appuyer sur un modèle de représentation des démarches ayant les caractéristiques suivantes :

- Une représentation modulaire des démarches de recherche d'information afin de favoriser leur partage et leur réutilisation,
- Une représentation des intentions (c'est-à-dire du pourquoi) de la recherche afin

de favoriser le transfert de connaissances des experts vers les néophytes et ainsi permettre à ces derniers de comprendre de façon globale un sujet sur lequel aucune source de données ne contient l'ensemble des informations pertinentes,

- Une représentation de la connaissance sur les démarches à plusieurs niveaux d'abstraction afin de prendre en considération les différents niveaux d'expertise des membres de la communauté.

Pour cela, l'approche originale que nous proposons de mettre en oeuvre est fondée sur l'adaptation au Web sémantique de la représentation intentionnelle d'un processus proposé dans (Rolland *et al.*, 1999; Rolland, 2007).

2.1 Le modèle de carte

D'après (Rolland *et al.*, 1999; Rolland, 2007), une carte est un modèle de processus dans lequel un ordonnancement non déterministe d'intentions et de stratégies de réalisation de ces intentions est représenté. Dans notre cas, nous nous concentrons sur des intentions et des stratégies de recherche. Une carte est un graphe nommé orienté ayant des intentions pour noeuds et des stratégies pour arcs entre les intentions. Une intention de recherche représente un but qui peut être atteint en suivant une stratégie de recherche. Une intention exprime ce qui est voulu (un état, un résultat) indépendamment de par qui, quand et où l'intention est réalisée. Deux intentions particulières sont distinguées : l'intention de début de démarche (*début*) et l'intention de fin de la démarche (*fin*). Une carte consiste donc en un ensemble de sections, chacune d'elle représentée par un triplet (intention source, stratégie, intention cible). Une stratégie représente la façon dont l'intention cible est réalisable à partir d'une intention source. Une carte contient un ensemble fini de chemins de l'intention *début* à l'intention *fin*, chacun décrivant une façon de satisfaire le but de la démarche de recherche d'information décrite.

La figure 1 montre un exemple de démarche pour rechercher des informations sur les bases de données relationnelles, par exemple dans le cadre de la conception d'un cours sur ce sujet dans une communauté d'enseignants. Pour mener à bien l'intention globale de la démarche (rechercher des ressources sur les bases de données relationnelles) il est recommandé de décomposer la recherche suivant les intentions présentées dans la figure 1. On peut remarquer que l'intention de recherche de ressources sur l'historique des bases de données relationnelles est facultative (il existe deux chemins de l'intention *début* à l'intention *fin*, l'un incluant cette intention, l'autre non) et que deux stratégies sont proposées pour réaliser l'intention de chercher des ressources sur le pilotage d'une base de données relationnelle depuis un langage de programmation.

Toujours d'après (Rolland *et al.*, 1999), à chaque section de la carte correspond une directive de réalisation d'intention (DRI) qui fournit des moyens opérationnels ou intentionnels de réaliser l'intention cible. Dans notre approche, une DRI opérationnelle correspond à l'exécution d'une requête SPARQL sur la mémoire de la communauté et une DRI intentionnelle est représentée par une carte définissant de façon plus détaillée (i.e. décomposant l'intention cible en sous-buts) la stratégie permettant de réaliser l'intention cible. La figure 2 montre un exemple de DRI intentionnelle pour la section mise en évidence dans la figure 1 et un exemple de DRI opérationnelle pour la section ayant l'intention *début* comme intention source dans la DRI intentionnelle de la figure 2.

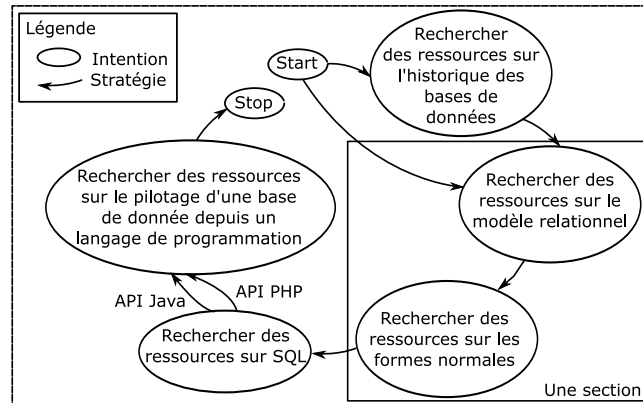


FIGURE 1 – Exemple de démarche

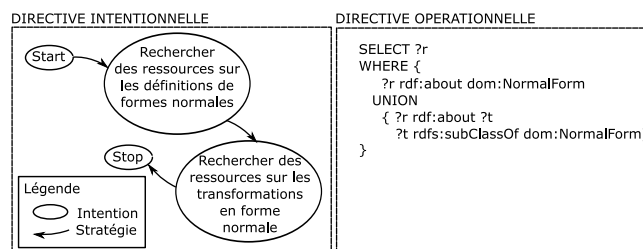


FIGURE 2 – Exemple de DRI

2.2 Intentions et stratégies

Pour permettre la réutilisation et le partage de requêtes de recherche d'information (i.e. DRI opérationnelles) et de démarches de recherche d'information (i.e. DRI intentionnelles), nous formalisons les notions de stratégie et d'intention. Pour cela, nous nous appuyons sur les travaux présentés dans (Prat, 1999), qui se sont déjà montrés pertinents pour formaliser des buts (Ralyte, 2001; Guzelian, 2007; Rolland, 2007). D'après (Prat, 1999), une intention est caractérisée par un verbe et des paramètres qui jouent un rôle particulier vis-à-vis du verbe. Parmi ces paramètres, il y a notamment l'objet sur lequel porte l'action décrite par le verbe. Dans (Ralyte, 2001; Rolland, 2007) d'autres paramètres sont proposés tels que le bénéficiaire du résultat de la réalisation de l'intention, l'objet à partir duquel l'action est réalisée ou une propriété qui devra être préservée durant la réalisation de l'intention. Ces paramètres permettent de formaliser la notion de stratégie. Dans l'exemple de la figure 1, les intentions sont spécifiées à l'aide du verbe *Search* et des objets du domaine, par exemple *Normal Form* et *Relational Model*; les stratégies à l'aide des paramètres *Java API* et *PHP API*.

3 Démarches sémantiques de recherche d'information

Nous considérons les démarches comme des ressources de la communauté et, à ce titre, nous les annotons afin de les indexer dans la mémoire collective de la communauté. Au-delà de la proposition d'un moyen alternatif d'organiser et d'accéder aux ressources d'une mémoire communautaire à travers la spécification de démarches décrivant des stratégies d'accès aux ressources pertinentes, l'annotation des démarches est un moyen de capitaliser la connaissance intrinsèque aux démarches elles-mêmes.

3.1 Une ontologie pour les démarches sémantiques de recherche d'information

Nous avons construit une ontologie RDFS pour annoter les démarches de recherche d'information qui rassemble les concepts et les relations du modèle de carte et ceux qui participent à la définition des notions d'intention et de stratégie telles que définies plus haut.

On retrouve comme classes principales de cette ontologie *Section*, *Intention*, *IntentionAchievementGuideline* qui représente une DRI, et *Resource* qui représente les ressources faisant l'objet du processus de recherche. Les intentions de début et de fin de démarche apparaissent comme sous-classes de la classe *Intention*. Une section est constituée d'une intention source, d'une intention cible et d'une stratégie (propriétés *hasTarget*, *hasSource* et *hasStrategy*). Des ressources Web sont dynamiquement associées aux sections dont elles permettent de satisfaire les intentions à l'aide de la propriété *hasResource*. Chaque section est rendue opérationnelle par une DRI (propriété *operationalizedBy*). Une intention est spécifiée à l'aide d'un verbe et d'un objet (classes *Verb* et *Object* et des propriétés *hasVerb* et *hasObject*). La classe *IntentionAchievementGuideline* est spécialisée en deux sous-classes *Map* et *GenericQuery* qui traduisent respectivement les notions de DRI intentionnelle et de DRI opérationnelle introduites précédemment.

Nous n'exploitons pour l'instant qu'un seul verbe, correspondant à la classe *Search* qui est instance de la (méta-) classe *Verb*, et nous considérons les concepts du domaine d'application de la recherche d'information comme des instances de la (méta-) classe *Object*. Les sous-classes de la classe *Parameter* sont instanciées en différentes classes qui modélisent une forme de contexte dans les processus de recherche d'information. Nous distinguons les informations contextuelles qui dépendent du domaine des informations recherchées de celles qui en sont indépendantes. Par exemple, les classes *Neophyte* et *Expert*, instances de la (méta-) classe *Beneficiary* sont donc toutes deux indépendantes du domaine : elles indiquent à quel type de membre le résultat de l'exécution de la DRI associée à une section de démarche est dédié. De même, les classes *DetailedDescription* et *ShortDescription*, instances de la classe *Quality*, sont indépendantes du domaine : elles indiquent si le résultat de l'exécution de la DRI associée à une section doit être constitué de ressources décrivant de façon détaillée ou non les informations résultat de la recherche. Au contraire, les classes *Java API* et *PHP API* instances de la classe *Manner* sont des exemples de concepts dépendants du domaine.

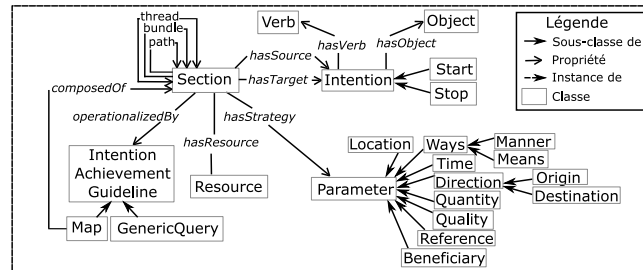


FIGURE 3 – Une ontologie pour les démarches de recherche d'information

Nous travaillons actuellement à la construction d'une ontologie indépendante du domaine d'application pour enrichir les extensions des concepts *Verb* et *Parameter*.

La figure 3 montre l'ontologie que nous avons construite pour annoter les démarches de recherche d'information.

3.2 Représentation des démarches de recherche d'information

Nous représentons les démarches de recherche d'information par des annotations RDF qui reposent sur l'ontologie décrite ci-dessus. Nous exploitons les capacités de raisonnement qu'offre le modèle RDF pour organiser les démarches de recherche d'information et pour les retrouver afin de les réutiliser.

Sur la carte décrite dans la figure 2, la section ayant pour source l'intention intitulée «Rechercher des ressources sur les définitions de formes normales» et pour cible celle intitulée «Rechercher des ressources sur les transformations en forme normale» est représentée en RDF de la façon suivante où *map* correspond au namespace de l'ontologie pour les démarches de recherche d'information et *dom* celui de l'ontologie de domaine :

```
<rdf:RDF xmlns:rdf="..." xmlns:map="..." xmlns:dom="...">
  <map:Section>
    <map:hasSource>
      <map:Intention rdf:nodeID="ii">
        <map:hasVerb rdf:resource="&dom;Search"/>
        <map:hasObject rdf:resource="&dom;NFDefinition"/>
      </map:Intention>
    </map:hasSource>
    <map:hasTarget>
      <map:Intention rdf:nodeID="ij">
        <map:hasVerb rdf:resource="&dom;Search"/>
        <map:hasObject rdf:resource="&dom;NFRule"/>
      </map:Intention>
    </map:hasTarget>
  </map:Section>
</rdf:RDF>
```


3.3 Partage et réutilisation de démarches

3.3.1 Des fragments de démarches

Une DRI et la section à laquelle elle est associée forment ce que nous appelons un fragment de démarche de recherche d'information. La définition de la section constitue la signature du fragment. Elle précise l'intention source, l'intention cible et la stratégie du fragment (seule la spécification de l'intention cible est obligatoire). Les intentions sont spécifiées à l'aide d'instances de la classe *Object* et d'instances de la classe *Verb*. La stratégie est spécifiée à l'aide d'instances de la classe *Parameter*.

La DRI constitue le corps du fragment de démarche. Elle est soit intentionnelle soit opérationnelle. Une DRI intentionnelle est une carte permettant de satisfaire l'intention cible considérée dans la signature du fragment, en partant éventuellement de l'intention source si elle est présente dans la signature du fragment et en suivant une stratégie éventuellement décrite dans la signature du fragment. Une DRI opérationnelle est une requête permettant de retrouver les ressources pertinentes dans la mémoire de la communauté. Les fragments correspondants aux DRI présentées dans la figure 2 sont représentés dans la figure 4.

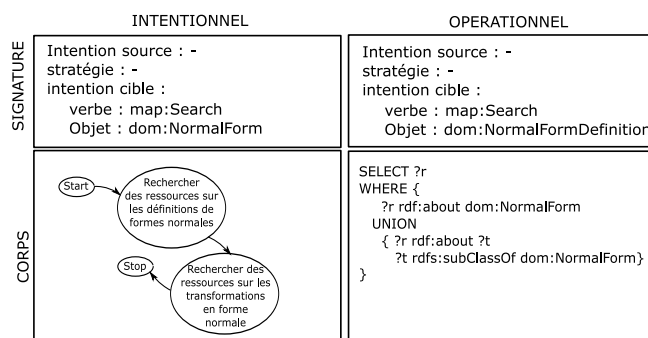


FIGURE 4 – Exemples de fragments et de règles

Lorsqu'un membre expert de la communauté souhaite conserver dans la mémoire de la communauté une recherche d'information, il définit sa démarche sous forme d'un ensemble de sections reliées les unes aux autres et chacune opérationnalisée à l'aide d'au moins un fragment, nouvellement créé pour l'occasion ou réutilisé parmi les fragments déjà présents dans la mémoire de la communauté. Les DRI intentionnelles des nouveaux fragments devront à leur tour être rendues opérationnelles jusqu'à proposer des DRI opérationnelles pour tous les sous-buts de la démarche.

Un fragment de démarche peut être réutilisé d'une démarche à l'autre si ces démarches partagent une même section ou des éléments de section. En effet, deux démarches incluant la réalisation d'une même intention peuvent être partiellement opérationnalisées par un même fragment dont la signature ne porte que sur la réalisation de l'intention cible (et ne contraint ni la situation de départ ni la stratégie de réalisation de l'intention).

La façon dont nous concevons la réutilisation de démarches de recherche d'infor-

mation est fondée sur la connexion dynamique de différents fragments pour rendre opérationnelle une démarche dans son ensemble en combinant les fragments de démarche d'après leur signature. Par exemple, un fragment ayant une intention cible *I* peut se combiner avec tout fragment ayant pour intention source *I*. Plusieurs fragments ayant *IS* pour intention source, *S* pour stratégie et *IC* pour intention cible peuvent être associés à une même section de démarche de recherche d'information et ainsi constituer autant de moyens différents de rendre opérationnelle la section de démarche considérée, c'est-à-dire de trouver un ensemble de ressources pertinentes dans la mémoire sémantique de la communauté.

3.3.2 Des fragments aux règles

Dans notre approche, nous implémentons chaque fragment de démarche sous forme d'une règle dont la conclusion correspond à la signature du fragment (une section de carte) et dont la prémisse correspond au corps du fragment (c'est-à-dire soit une requête SPARQL dans le cas d'une DRI opérationnelle soit une carte dans le cas d'une DRI intentionnelle). Nous distinguons les règles concrètes dont la prémisse est constituée d'une requête SPARQL des règles abstraites dont la prémisse est constituée d'une carte. Le langage SPARQL constitue un cadre unifié pour représenter règles concrètes et règles abstraites en reposant sur la forme de requête CONSTRUCT-WHERE. La clause WHERE d'une telle requête peut être vue comme la prémisse d'une règle et la clause CONSTRUCT comme sa conclusion. Une telle requête construit des graphes RDF en substituant aux variables de sa clause CONSTRUCT les valeurs qui satisfont sa clause WHERE (retrouvées en recherchant les appariements possibles de sa clause WHERE avec les données RDF interrogées). Ainsi, nous représentons un fragment de démarche par une requête SPARQL dont la clause CONSTRUCT est un patron de graphe permettant de construire la représentation RDF de la section de la carte considérée et dont la clause WHERE est un patron de graphe qui représente soit une carte (règle abstraite) soit un critère de recherche de ressources pertinentes (règle concrète). Dans le cas d'une règle abstraite, la clause WHERE est un patron de graphe représentant une sous-carte permettant la réalisation de l'intention cible du fragment. Dans le cas d'une règle concrète, la clause WHERE est un graphe qui permet de retrouver les ressources pertinentes, c'est-à-dire celles avec les annotations RDF desquelles il existe des appariements avec ce graphe requête. Les règles correspondant aux fragments présentés dans la figure 4 sont représentées dans la figure 5.

Une section de carte peut être implémentée par différentes requêtes, concrètes ou abstraites, afin de proposer différentes possibilités de recherche à différents niveaux d'abstraction. Ces requêtes partagent la même clause CONSTRUCT (fragments de même signature) et diffèrent par le contenu de leur clause WHERE.

La mise en oeuvre d'une démarche de recherche d'information, i.e. la combinaison de fragments de démarche en une démarche globale, est réalisée en appliquant les règles qui implémentent les fragments de démarche en chaînage arrière. Nous nous appuyons sur le moteur sémantique CORESE¹ (Corby *et al.*, 2006) à la fois pour le chaînage arrière sur la base de requêtes SPARQL représentant des règles et pour trouver les réponses à

1. <http://www-sop.inria.fr/edelweiss/software/corese/>

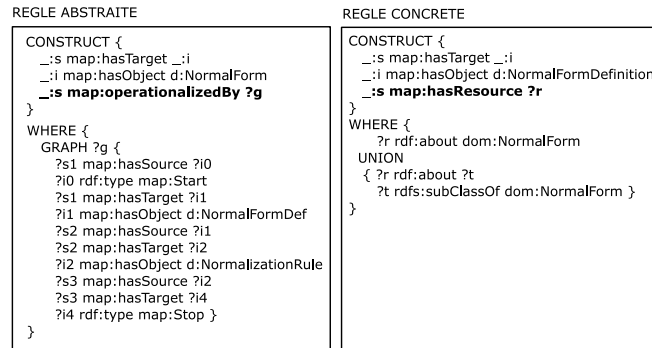


FIGURE 5 – Exemples de fragments et de règles

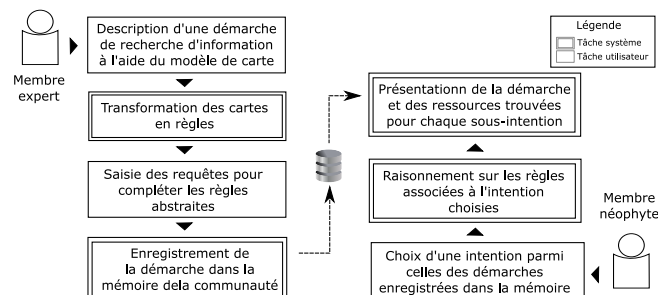


FIGURE 6 – Processus de définition et de mise en oeuvre de démarches

une requête dans la base d'annotations RDF des ressources de la mémoire sémantique de la communauté. Soulignons que seules les requêtes sont capitalisées dans la mémoire de la communauté et non les cartes. Ces dernières sont créées dynamiquement tout au long du processus de chaînage arrière, comme des sous-buts temporaires, jusqu'à ce que toutes les ressources dont les annotations correspondent à des sous-buts et donc à l'intention globale de la démarche de recherche d'information soient retrouvées. Un membre de la communauté qui souhaite trouver une démarche de recherche d'information sur un sujet donné doit en fait rechercher la règle qui correspond à l'intention de recherche qu'il souhaite satisfaire (intention globale) et bénéficie de toutes les règles et de toutes les annotations de ressources présentes dans la mémoire sémantique de la communauté au moment de la mise en oeuvre de la démarche. La mémoire de la communauté évolue au cours du temps et ainsi l'ensemble des ressources retournées peut également varier dans le temps. En d'autres termes, l'instantiation de la démarche de recherche d'information dans son ensemble est réalisée au moment de son exécution et dépend donc des ressources disponibles dans la mémoire collective à ce moment-là. Le processus de définition d'une démarche par un expert et le processus de mise en oeuvre d'une démarche par un néophytes sont schématisés dans la figure 6.

4 Conclusion

Dans cet article nous avons présenté une approche, fondée sur les techniques et modèles du Web sémantique, qui a pour but de capitaliser, partager et réutiliser des requêtes et des démarches de recherche d'information. En proposant une modélisation intentionnelle et sémantique des démarches de recherche d'information, notre but est de capitaliser de la connaissance sur les pratiques en recherche d'information au sein d'une communauté et cela au travers de séquences structurées de tâches de recherche. Pour cela, en partant d'un formalisme de représentation des processus guidé par les intentions, nous avons proposé une ontologie permettant d'annoter des démarches de recherche d'information et nous avons opérationnalisé les directives correspondant aux fragments de ces démarches à l'aide de règles implémentées par des requêtes SPARQL. Les démarches de recherche d'information sont mises en oeuvre à l'aide d'un mécanisme de chaînage arrière appliqué sur la base de règles et sur les annotations RDF des ressources de la communauté.

Cinq perspectives se dessinent à l'issue de ce travail :

- La proposition d'une ontologie de concepts indépendants du domaine d'application pour enrichir l'ensemble des classes instances de *Verb* et *Parameter* dans l'ontologie des démarches ;
- La proposition d'un modèle de requête générique permettant la représentation de requêtes récurrentes dans les fragments de démarche ;
- La prise en compte des profils des membres de la communauté et du contexte de mise en oeuvre de la démarche afin d'affiner le processus de sélection des fragments de démarche lors de l'opérationnalisation d'une démarche de recherche d'information ;
- La proposition d'une ontologie et d'un modèle pour annoter les démarches et ainsi faciliter leur partage au sein de la communauté, notamment en fonction des profils des membres et de leurs contextes de travail.
- Le développement d'un outil pour mettre en oeuvre cette approche et la proposition d'interfaces idoines pour aider les membres de la communauté à naviguer dans la mémoire sémantique dédiée aux stratégies de recherche.

Ce travail est le point de départ du projet DESIR (COLOR INRIA) qui démarre cette année en partenariat avec deux équipes de l'INRA et qui vise à l'explicitation et la capitalisation des processus de recherche d'information d'agronomes et généticiens sur différentes bases de données hétérogènes.

Références

- BHAVNANI S., BICHAKJIAN C., JOHNSON T., LITTLE R., PECK F., SCHWARTZ J. & STRECHER V. (2003). Strategy hubs : Next-generation domain portals with search procedures. In G. COCKTON & P. KORHONEN, Eds., *ACM Conference on Human Factors in Computing Systems*, p. 393–400, Florida, USA : ACM.
- BUFFEREAU B., DUCHET P. & PICOUEP P. (2003). Generating guided tours to facilitate learning from a set of indexed resources. In *IEEE International Conference on*

- Advanced Learning Technologies (ICALT)*, p. 492, Athens, Greece : IEEE Computer Society.
- CHERFI H., CORBY O., FARON-ZUCKER C. & KHELIF K. (2008). Semantic annotation of texts with RDF graph contexts. In P. W. EKLUND & O. HAEMMERLÉ, Eds., *International Conference on Conceptual Structures (ICCS'2008)*, p. 75–82, Krakow, Poland : CEUR-WS.org.
- CORBY O., DIENG-KUNTZ R., FARON-ZUCKER C. & GANDON F. (2006). Searching the semantic web : Approximate query processing based on ontologies. *IEEE Intelligent Systems Journal*, **21**(1).
- DEHORS S. & FARON-ZUCKER C. (2006). Qbls : A semantic web based learning system. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA)*, Orlando, FL, USA.
- EWOKHUB CONSORTIUM (2008). Semantic hubs for geographical projects. In CEUR-WS.ORG, Ed., *Semantic Metadata Management and Applications (SeMMA), workshop at ESWC*, p. 3–17, Teneriffe, Spain : Khalid Belhajjame and Mathieu d'Aquin and Peter Haase and Paolo Missier.
- GUZELIAN G. (2007). *Modelisation et specification de composants réutilisables pour la conception de systèmes d'information*. PhD thesis, Université Aix Marseille.
- KEFI-KHELIF L., DEMARCHEZ M. & COLLARD M. (2008). A knowledge base approach for genomics data analysis. In *International Conference on Semantic Systems, I-Semantics 2008*, Graz, Austria.
- NURCAN S. & EDME M. (2005). Intention-driven modeling for flexible workflow applications. *Software Process : Improvement and Practice*, **10**(4), 363–377.
- PRAT N. (1999). *Réutilisation de la trace par apprentissage dans un environnement pour l'ingénierie des processus*. PhD thesis, Université Paris I - Sorbonne.
- RALYTE J. (2001). *Ingénierie des méthodes à base de composants*. PhD thesis, Université Paris I - Sorbonne.
- ROLLAND C. (2007). *Conceptual Modelling in Information Systems Engineering*, chapter Capturing System Intentionality with Maps. Springer-Verlag.
- ROLLAND C., PRAKASH N. & BENJAMEN A. (1999). A multi-model view of process modelling. *Requirements Engineering*, **4**(4), 169–187.
- YESSAD A., FARON-ZUCKER C., DIENG-KUNTZ R. & LASKRI M. (2008). Ontology-driven adaptive course generation for web-based education. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA)*, Vienna, Austria.
- YURCHYSHYNA A., FARON-ZUCKER C., MIRBEL I., SALL B., LE THANH N. & ZARLI A. (2008). Une approche ontologique pour formaliser la connaissance experte dans le modèle du contrôle de conformité en construction. In Y. T. YANNICK PRIÉ, Ed., *19ième journées francophones d'ingénierie des connaissances*, Nancy, France : Capaudes Editions.